On Polynomial Approximations Over $\mathbb{Z}/2^k\mathbb{Z}$[*][†]

Abhishek Bhrushundi[1], Prahladh Harsha[2], and
Srikanth Srinivasan[3]

1 **Department of Computer Science, Rutgers University, New Brunswick, USA**
abhishek.bhr@cs.rutgers.edu

2 **Tata Institute of Fundamental Research, Mumbai, India**
prahladh@tifr.res.in

3 **Department of Mathematics, IIT Bombay, Bombay, India**
srikanth@math.iitb.ac.in

Abstract

We study approximation of Boolean functions by low-degree polynomials over the ring $\mathbb{Z}/2^k\mathbb{Z}$. More precisely, given a Boolean function $F : \{0,1\}^n \to \{0,1\}$, define its k-lift to be $F_k : \{0,1\}^n \to \{0, 2^{k-1}\}$ by $F_k(x) = 2^{k-F(x)} \pmod{2^k}$. We consider the fractional agreement (which we refer to as $\gamma_{d,k}(F)$) of F_k with degree d polynomials from $\mathbb{Z}/2^k\mathbb{Z}[x_1, \ldots, x_n]$.

Our results are the following:

- Increasing k can help: We observe that as k increases, $\gamma_{d,k}(F)$ cannot decrease. We give two kinds of examples where $\gamma_{d,k}(F)$ actually increases. The first is an infinite family of functions F such that $\gamma_{2d,2}(F) - \gamma_{3d-1,1}(F) \geq \Omega(1)$. The second is an infinite family of functions F such that $\gamma_{d,1}(F) \leq \frac{1}{2} + o(1)$ – as small as possible – but $\gamma_{d,3}(F) \geq \frac{1}{2} + \Omega(1)$.
- Increasing k doesn't always help: Adapting a proof of Green [*Comput. Complexity*, 9(1):16–38, 2000], we show that irrespective of the value of k, the Majority function Maj_n satisfies

$$\gamma_{d,k}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}}.$$

  In other words, polynomials over $\mathbb{Z}/2^k\mathbb{Z}$ for large k do not approximate the majority function any better than polynomials over $\mathbb{Z}/2\mathbb{Z}$.

We observe that the model we study subsumes the model of *non-classical polynomials* in the sense that proving bounds in our model implies bounds on the agreement of non-classical polynomials with Boolean functions. In particular, our results answer questions raised by Bhowmick and Lovett [*In Proc. 30th Computational Complexity Conf., pages 72—87, 2015*] that ask whether non-classical polynomials approximate Boolean functions better than classical polynomials of the same degree.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Polynomials over rings, Approximation by polynomials, Boolean functions, Non-classical polynomials

Digital Object Identifier 10.4230/LIPIcs.STACS.2017.12

[*] A full version of the paper is available at https://arxiv.org/abs/1701.06268.

[†] Research supported in part by UGC-ISF grant 1399/4. Part of the work done when the first author was visiting TIFR.

34th Symposium on Theoretical Aspects of Computer Science (STACS 2017).
Editors: Heribert Vollmer and Brigitte Vallée; Article No. 12; pp. 12:1–12:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Many lower bound results in circuit complexity are proved by showing that any small sized circuit in a given circuit class can be approximated by a function from a simple computational model (e.g., small depth circuits by low-degree polynomials) and subsequently showing that this is not possible for some suitable "hard" function.

A classic case in point is the work of Razborov [12] which shows lower bounds for $\mathrm{AC}^0[\oplus]$, the class of constant depth circuits made up of AND, OR and $\oplus$ gates. Razborov shows that any small $\mathrm{AC}^0[\oplus]$ circuit C can be well approximated by a low-degree multivariate polynomial $Q(x_1, \ldots, x_n) \in \mathbb{F}_2[x_1, \ldots, x_n]$ in the sense that

$$\Pr_{x\sim\{0,1\}^n} [Q(x) \neq C(x)] = o(1).$$

The next step in the proof is to show that the hard function, on the other hand, does not have any such approximation. Razborov does this for a suitable symmetric function, Smolensky [13] for the MOD_q function (for constant odd q), and Szegedy [15] and Smolensky [14] for the Majority function Maj_n on n bits.

Given the importance of the above lower bound, polynomial approximations in other domains and metrics have been intensely investigated and have resulted in interesting combinatorial constructions and error-correcting codes [8, 4], learning algorithms [11, 9] and more recently in the design of algorithms for combinatorial problems [17, 1] as well.

To describe the model of polynomial approximation considered in this paper, we first recall the Razborov [12] model of polynomial approximation. Given a Boolean function $F : \{0,1\}^n \to \{0,1\}$ and degree $d \leq n$, Razborov considers the largest γ such that there is a degree d polynomial $Q \in \mathbb{F}_2[x_1, \ldots, x_n]$ that has agreement at least γ with F (i.e., $\Pr_x[Q(x) = F(x)] \geq \gamma$). Call this $\gamma_d(F)$. In this notation, Szegedy [15] and Smolensky's [14] results for the Majority function can be succinctly stated as

$$\gamma_d(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}}.$$

We consider a generalization of the above model to rings $\mathbb{Z}/2^k\mathbb{Z}$ in the following simple manner. To begin with, we consider the ring $\mathbb{Z}/4\mathbb{Z}$. Given a Boolean function F, let $F_2 : \{0,1\}^n \to \{0,2\} \subseteq \mathbb{Z}/4\mathbb{Z}$ be the 2-lift of F defined as $F_2(x) := 2^{2-F(x)}$ (i.e., $F_2(x) := 0$ if $F(x) = 0$ and $F_2(x) := 2$ otherwise). Once again, we can define $\gamma_{d,2}(F)$ to be the largest γ such that there exists a degree d polynomial $Q_2 \in \mathbb{Z}/4\mathbb{Z}[x_1, \ldots, x_n]$ that has agreement γ with F_2. Note that $\gamma_{d,2}(F) \geq \gamma_d(F)$ since if, for instance, $Q(x) = x_1x_2 + x_3 \in \mathbb{F}_2[x_1, \ldots, x_n]$ has agreement γ with F, then $Q_2 := 2(x_1x_2 + x_3) \in \mathbb{Z}/4\mathbb{Z}[x_1, \ldots, x_n]$ also has the same agreement γ with F_2. Hence, proving upper bounds for $\gamma_{d,2}(F)$ is at least as hard as proving upper bounds for $\gamma_d(F)$.

More generally, we can extend these definitions to $\gamma_{d,k}(F)$, the agreement of F_k, the k-lift of F, defined as $F_k(x) = 2^{k-F(x)} \mod 2^k$, with degree d polynomials from $\mathbb{Z}/2^k\mathbb{Z}[x_1, \ldots, x_n]$. It is not hard to show that $\gamma_{d,k+1}(F) \geq \gamma_{d,k}(F)$ and hence as k increases, the problem of proving upper bounds on $\gamma_{d,k}(F)$ can only get harder.

Our motivation for this model comes from a recent work of Bhowmick and Lovett [3], who study the maximum agreement between *non-classical polynomials* of degree d and a Boolean function F, which is similar to $\gamma_{d,d}(F)$ (see Section 5 for an exact translation between the above model and non-classical polynomials). In particular, non-classical polynomials of degree d can be considered as a subset of the degree d polynomials in $\mathbb{Z}/2^d\mathbb{Z}[x_1, \ldots, x_n]$.

With respect to correlation[1], Bhowmick and Lovett showed that there exist non-classical polynomials (and hence polynomials in $\mathbb{Z}/2^d\mathbb{Z}[x_1,\ldots,x_n]$) of logarithmic degree that have very good correlation with the Maj_n function. With respect to agreement, they show that low-degree non-classical polynomials can only have *small* agreement with the Majority function. Their results stated in our language, imply that

$$\gamma_{d,d}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d \cdot 2^d)}{\sqrt{n}}.$$

In particular, if $d = \Omega(\log n)$, this result unfortunately does not give any non-trivial bound on the maximum agreement between non-classical polynomials of degree d and the Maj_n function. Bhowmick and Lovett, however, conjectured that this result could be improved and left open the question of whether non-classical polynomials of degree d can do any better than classical polynomials of the same degree in approximating the Majority function. More generally, they informally conjectured that although non-classical polynomials achieve better correlation with Boolean functions than their classical counterparts, they possibly do not approximate Boolean functions any better than classical polynomials. Our work stems from trying to answer these questions.

1.1 Our results

We prove the following results about agreement of Boolean functions with polynomials over the ring $\mathbb{Z}/2^k\mathbb{Z}$:

1. We explore whether there exist Boolean functions for which agreement can increase by increasing k. In particular, do there exist Boolean F such that $\gamma_{d,k}(F) > \gamma_{d,1}(F)$?
   It is not hard to show that this is impossible for $d = 1$. Further, it can be shown that if $\gamma_{d,k}(F) > 1 - \frac{1}{2^d}$, then $\gamma_{d,k(}F) = \gamma_{d,1}(F)$. Keeping this in mind, the first place where we can expect larger k to show better agreement is $\gamma_{2,2}$ vs. $\gamma_{2,1}$. Our first result shows that there are indeed separating examples in the regime.
   (a) Fix $d \in \mathbb{N}$ to be any power of 2. For infinitely many n, there exists a Boolean function $F : \{0,1\}^n \to \{0,1\}$ such that $\gamma_{3d-1,1}(F) \leq 5/8 + o(1)$ but $\gamma_{2d,2}(F) \geq 3/4$.
   Note that since F is Boolean, $\gamma_{d,k}(F) \geq 1/2$ for any d, k. We then ask if there exist Boolean functions F such that $\gamma_{d,1}(F)$ is more or less the trivial bound of 1/2, while $\gamma_{d',k}(F)$ is significantly larger for $d' \leq d$ and some $k > 1$. In this context, we show the following result.
   (b) Fix any $\ell \geq 2$. For large enough n, there is a Boolean function $F : \{0,1\}^n \to \{0,1\}$ such that $\gamma_{2^\ell-1,1}(F) \leq 1/2 + o(1)$ but $\gamma_{d,3}(F) \geq 9/16 - o(1)$, for $d = 2^{\ell-1} + 2^{\ell-2} \leq 2^\ell - 1$.
2. We show that for Maj_n, the majority function on n bits, and any $d, k \in \mathbb{Z}^+$,

$$\gamma_{d,k}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}},$$ [2]

   by adapting a proof due to Green [7] of a result on the approximability of the parity function by low-degree polynomials over the ring $\mathbb{Z}/p^k\mathbb{Z}$ for prime $p \neq 2$.

[1] The correlation between $F, G : \{0,1\}^n \to \mathbb{Z}/2^k\mathbb{Z}$ is defined to be $\mathbf{E}_x[\omega^{F(x)-G(x)}]$ where ω is the primitive 2^kth root of unity in $\mathbb{C}$. If F, G are $\{0, 2^{k-1}\}$-valued, then this quantity is exactly $2\gamma - 1$ where γ is the agreement between F and G. Otherwise, however, it does not measure agreement.

Coupled with the observation that the class of polynomials over rings $\mathbb{Z}/2^k\mathbb{Z}$ subsumes the class of non-classical polynomials, part (b) of the first result provides a counterexample to an informal conjecture of Bhowmick and Lovett [3] that, for any Boolean function F, non-classical polynomials of degree d do not approximate F any better than classical polynomials of the same degree, and the second result confirms their conjecture that non-classical polynomials do not approximate the Majority function any better than classical polynomials.

1.2 Organization

We start with some preliminaries in Section 2. In Section 3, we show some separation results. Next, in Section 4, we prove upper bounds for $\gamma_{d,k}(\mathrm{Maj}_n)$. Finally, in Section 5, we discuss how our model relates to non-classical polynomials, answering questions raised by Bhowmick and Lovett.

2 Preliminaries

For $x \in \{0,1\}^n$, $|x|$ denotes the Hamming weight of x, and for $i \geq 0$, $|x|_i$ is the $(i+1)^{\text{th}}$ least significant bit of $|x|$ in base 2. For $d \in \mathbb{N}$, we use $\{0,1\}^n_{\leq d}$ (resp. $\{0,1\}^n_{=d}$) to denote the set of elements in $\{0,1\}^n$ of Hamming weight at most d (resp. exactly d). We use $\mathcal{F}_n$ to denote the collection of all Boolean functions defined on $\{0,1\}^n$.

2.1 Elementary symmetric polynomials

Recall that for $t \geq 1$, the elementary symmetric polynomial of degree t over $\mathbb{F}_2$, $S_t(x_1,\ldots,x_n)$, is defined as $S_t(x_1,\ldots,x_n) = \bigoplus_{1\leq a_1<\ldots<a_t\leq n} x_{a_1}\ldots x_{a_t}$. Here $\oplus$ denotes addition modulo two. This may be interpreted as

$$S_t(x_1,\ldots,x_n) = \binom{|x|}{t} \bmod 2. \tag{1}$$

A direct consequence of Lucas theorem (see, e.g., [10, Section 1.2.6, Ex. 10]) and Equation (1) is the following:

▶ **Lemma 2.1.** *For every $\ell \geq 0$, $S_{2^\ell}(x) = |x|_\ell$. More generally, $S_t(x) = \prod_i |x|_i$ where the product runs over all $i \geq 0$ such that the $(i+1)^{th}$ least significant bit of the binary expansion of t is* 1.

The following result follows from the work of Green and Tao [6, Theorem 11.3], who build upon the ideas of Alon and Beigel [2].

▶ **Theorem 2.2** (Alon-Beigel [2]). *Fix $\ell \geq 0$. Then, for every multilinear polynomial $P \in \mathbb{F}_2[x_1,\ldots,x_n]$ of degree at most $2^\ell - 1$, we have $\Pr_{x\sim\{0,1\}^n}[S_{2^\ell}(x) = P(x)] \leq 1/2 + o(1)$.*

Theorem 2.2 has a nice corollary:

▶ **Corollary 2.3.** *For every fixed $\ell \geq 0$, the functions $\{S_{2^i}(x)\}_{0\leq i\leq \ell}$ are almost balanced and almost uncorrelated, i.e.*

- $\forall\ 0 \leq i \leq \ell$, $|\Pr[S_{2^i}(x) = 0] - \Pr[S_{2^i}(x) = 1]| = o(1)$
- $\forall\ a_0,\ldots,a_\ell \in \{0,1\}$, $|\Pr[\bigwedge_{0\leq i\leq \ell} S_{2^i}(x) = a_i] - \frac{1}{2^{\ell+1}}| = o(1)$.

[2] The constant in the $O(\cdot)$ is an absolute constant.

Combining Corollary 2.3 with Lemma 2.1, we get another useful fact:

▶ **Lemma 2.4.** *Let x be uniformly distributed over $\{0,1\}^n$. Then, for every fixed $r \geq 1$, the random variables $\{|x|_i\}_{0 \leq i \leq r-1}$ are almost uniform and almost r-wise independent i.e.*

- $\forall\ 0 \leq i \leq r-1$, $|\Pr[|x|_i = 0] - \Pr[|x|_i = 1]| = o(1)$.
- $\forall\ (a_0, \ldots, a_{r-1}) \in \{0,1\}^r$, $|\Pr[(|x|_0, \ldots, |x|_{r-1}) = (a_0, \ldots, a_{r-1})] - \frac{1}{2^r}| = o(1)$.

2.2 Boolean functions and polynomials over $\mathbb{Z}/2^k\mathbb{Z}$

Given an $F \in \mathcal{F}_n$ and $k \geq 1$, we define the *k-lift of F* to be the function $F_k : \{0,1\}^n \to \mathbb{Z}/2^k\mathbb{Z}$ defined as follows. For any $x \in \{0,1\}^n$,

$$F_k(x) = \begin{cases} 0 & \text{if } F(x) = 0, \\ 2^{k-1} & \text{otherwise.} \end{cases}$$

For $d \in \mathbb{N}, k \geq 1$, $\mathcal{P}_{d,k}$ will denote the set of multilinear polynomials in $\mathbb{Z}/2^k\mathbb{Z}[x_1, \ldots, x_n]$ of degree at most d .

For functions $F, G : D \to R$ for some finite domain D and range R, define the *agreement of F and G* – denoted $\operatorname{agr}(F,G)$ – to be the fraction of inputs where they agree: i.e.,

$$\operatorname{agr}(F,G) = \Pr_{x \sim D}[F(x) = G(x)].$$

We will consider how well multilinear polynomials of a certain degree can approximate Boolean functions in the above sense. More precisely, for any Boolean function $F \in \mathcal{F}_n$, we define

$$\gamma_{d,k}(F) = \max_{Q \in \mathcal{P}_{d,k}} \operatorname{agr}(F_k, Q).$$

Following [5], we call a set $I \subseteq \{0,1\}^n$ an *interpolating set* for $\mathcal{P}_{d,k}$ if the only polynomial $P \in \mathcal{P}_{d,k}$ that vanishes at all points in I is zero everywhere. Formally, for any $P \in \mathcal{P}_{d,k}$,

$$(\forall x \in I \quad P(x) = 0) \Rightarrow (\forall y \in \{0,1\}^n \quad P(y) = 0).$$

We now state a number of standard facts regarding Boolean functions and multilinear polynomials over $\mathbb{Z}/2^k\mathbb{Z}$. The proofs are either easy or well-known, and appear in the full version of the paper (see `https://arxiv.org/abs/1701.06268`).

Unless mentioned otherwise, let n, d, k be any integers satisfying $n \geq 1, d \geq 0, k \geq 1$.

▶ **Lemma 2.5.** *Any polynomial $Q \in \mathcal{P}_{d,k}$ satisfies the following:*

1. *If Q is non-zero, then $\Pr_{x \sim \{0,1\}^n}[Q(x) \neq 0] \geq \frac{1}{2^d}$.*
2. *Q is the zero polynomial iff $Q(x) = 0$ for all $x \in \{0,1\}^n$.*
3. (Möbius Inversion) *Say $Q(x) = \sum_{|S| \leq d} c_S x_S$, where $c_S \in \mathbb{Z}/2^k\mathbb{Z}$ and x_S denotes $\prod_{i \in S} x_i$. Then, $c_S = \sum_{T \subseteq S} (-1)^{|S|-|T|} Q(1_T)$ where $1_T \in \{0,1\}^n$ is the characteristic vector of T.*
4. ($\{0,1\}^n_{\leq d}$ is an interpolating set) *Q vanishes at all points in $\{0,1\}^n$ iff Q vanishes at all points of $\{0,1\}^n_{\leq d}$. By shifting the origin to any point of $\{0,1\}^n$, the same is true of* any *Hamming ball of radius d in $\{0,1\}^n$.*

▶ **Lemma 2.6.** *Fix any $F \in \mathcal{F}_n$.*

1. $\gamma_{d,k}(F) \geq \frac{1}{2}$.
2. $\gamma_{d,k+1}(F) \geq \gamma_{d,k}(F)$.
3. $\gamma_{d,k}(F) > 1 - \frac{1}{2^d} \Rightarrow \gamma_{d,k}(F) = \gamma_{d,1}(F)$.
4. $\gamma_{1,k}(F) = \gamma_{1,1}(F)$.

3 Some separation results

3.1 A separation at $k = 2$

Let $d \in \mathbb{N}$ be any power of 2. In this section, we show that there are functions F for which $\gamma_{2d,2}(F) > \gamma_{3d-1,1}(F)$.

▶ **Theorem 3.1.** *For large enough n, there exists a function $F \in \mathcal{F}_{2n}$ such that $\gamma_{2d,2}(F) \geq \frac{3}{4} - o(1)$ but $\gamma_{3d-1,1}(F) \leq \frac{5}{8} + o(1)$.*

In particular, we see that $\gamma_{2,2}(F) > \gamma_{2,1}(F)$. This result is notable, since it shows that there is a separation at the first place where it is possible to have one (Recall that $\gamma_{1,k}(F) = \gamma_{1,1}(F)$ for any $F \in \mathcal{F}_n$ by Lemma 2.6).

Let us begin the proof of Theorem 3.1. We first define a family of Boolean functions on $\{0,1\}^{2n}$. We denote the $2n$ variables by $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. We use $\binom{|x|}{d}$ to denote the dth elementary symmetric polynomial from the ring $\mathbb{Z}/4\mathbb{Z}[x_1, \ldots, x_n]$, i.e., $\binom{|x|}{d} = \sum_{S \in \binom{[n]}{d}} \prod_{i \in S} x_i$.[3]

The following theorem due to Kummer (see, e.g., [10, Section 1.2.6, Ex. 11]) determines the largest power of a prime that divides a binomial coefficient.

▶ **Theorem 3.2** (Kummer). *Let p be a prime and $N, M \in \mathbb{N}$ such that $N \geq M$. Suppose r is the largest integer such that $p^r \mid \binom{N}{M}$. Then r is equal to the number of borrows required when subtracting M from N in base p.*

We will need the following easy corollary of Kummer's theorem.

▶ **Corollary 3.3.** *Let d be a power of 2. Then, for $N \geq d$, the highest power of 2 dividing $\binom{N}{d}$ is equal to the highest power of 2 dividing $\lfloor \frac{N}{d} \rfloor$.*

Let $S = \{(x, y) \mid \binom{|x|}{d}, \binom{|y|}{d} \equiv 1 \pmod 2\}$. Given any function $H : \{0,1\}^{2n} \to \{0,1\}$, we define the Boolean function $F_H(x_1, \ldots, x_n, y_1, \ldots, y_n)$ as follows:

$$F_H(x,y) = \begin{cases} 0 & \text{if } \binom{|x|}{d} \cdot \binom{|y|}{d} \equiv 0 \pmod 4, \\ 1 & \text{if } \binom{|x|}{d} \cdot \binom{|y|}{d} \equiv 2 \pmod 4, \\ H(x,y) & \text{otherwise.} \end{cases}$$

Define $P(x,y) = \binom{|x|}{d} \cdot \binom{|y|}{d} \in \mathbb{Z}/4\mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_n]$. Note that $F_H(x,y)$ is defined so that its 2-lift agrees with $P(x,y)$ on points (x,y) where $P(x,y) \in \{0,2\}$. Also Corollary 3.3 implies that the following is an alternate equivalent definition of F_H in terms of elementary symmetric polynomials modulo 2.

$$F_H(x,y) = \begin{cases} 0 & \text{if } S_d(x) = S_d(y) = 0, \\ S_{2d}(y) & \text{if } S_d(x) = 1 \text{ and } S_d(y) = 0, \\ S_{2d}(x) & \text{if } S_d(x) = 0 \text{ and } S_d(y) = 1, \\ H(x,y) & \text{otherwise.} \end{cases} \tag{2}$$

First of all, let us note that for any choice of H, we have:

▶ **Lemma 3.4.** $\gamma_{2d,2}(F_H) \geq \frac{3}{4} - o(1)$.

[3] We distinguish between $\binom{|x|}{d}$ and $S_d(x)$ since the former is from $\mathbb{Z}/4\mathbb{Z}[x_1, \ldots, x_n]$ and latter a polynomial in $\mathbb{F}_2[x_1, \ldots, x_n]$.

Proof. Consider the polynomial $P(x,y) \in \mathcal{P}_{2d,2}$ defined above. From Equation (2), it follows that the probability that $P(x,y) \neq F_{H,2}(x,y)$[4] is less than or equal to the probability that $S_d(x) = S_d(y) = 1$, which is $\frac{1}{4} + o(1)$ by Corollary 2.3. This gives the claim. ◀

The main lemma is the following.

▶ **Lemma 3.5.** *Say $H : \{0,1\}^{2n} \to \{0,1\}$ is chosen uniformly at random. Then,*

$$\Pr_H \left[\gamma_{3d-1,1}(F_H) > \frac{5}{8} + o(1) \right] = o(1).$$

This will prove Theorem 3.1.

The outline of the proof of the above lemma is as follows. Fix any polynomial $Q \in \mathbb{F}_2[x_1,\ldots,x_n,y_1,\ldots,y_n]$ of degree at most $3d-1$. We need to show that $\mathrm{agr}(F_H,Q) \leq \frac{5}{8}+o(1)$. The function $H : \{0,1\}^{2n} \to \{0,1\}$ we choose will be a *random* function, which ensures that any Q cannot agree with H on significantly more than half the inputs in S. For inputs outside S, we need a more involved argument, following Alon and Beigel [2]. We show that for any Q we can find somewhat large sets I and J of x and y variables respectively such that when we set the variables outside $I \cup J$, we obtain a polynomial that is symmetric in the variables of $I \cup J$. This is a Ramsey theoretic argument ála Alon-Beigel [2].

Following this argument, we only need to prove the agreement upper bound for Q that is symmetric in x and y variables. This can be done by reduction to a constant-sized problem. A careful computation to solve the constant-sized problem finishes the proof.

The complete technical details of the proof of Lemma 3.5 appear in the full version of the paper.

3.2 Symmetric functions as separating examples

We know from Theorem 2.2 that, for every fixed $\ell \geq 2$, $\gamma_{2^\ell-1,1}(S_{2^\ell}) \leq \frac{1}{2} + o(1)$. In contrast, the main result of this section shows that

▶ **Theorem 3.6.** *For every fixed $\ell \geq 2$, $\gamma_{d,3}(S_{2^\ell}) \geq \frac{9}{16} - o(1)$, where $d = 2^{\ell-1} + 2^{\ell-2}$.*

Notice that $2^{\ell-1} + 2^{\ell-2} \leq 2^\ell - 1$ for $\ell \geq 2$. This implies that, for $\ell \geq 2$, $S_{2^\ell}(x)$ is an example of a function F for which there exist $k, d \in \mathbb{N}$ such that $\gamma_{d,1}(F) \leq \frac{1}{2} + o(1)$ but $\gamma_{d',k}(F) \geq \frac{1}{2} + \Omega(1)$ for some $d' \leq d$.

Proof of Theorem 3.6. Lemma 2.1 from Section 2 tells us that $S_{2^\ell}(x) = |x|_\ell$. Thus, $S_{2^\ell,3}(x) \in \mathbb{Z}/8\mathbb{Z}[x_1,\ldots,x_n]$, the 3-lift of $S_{2^\ell}(x)$, is given by

$$S_{2^\ell,3}(x) = \begin{cases} 4 & \text{if } |x|_\ell = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Fix d to be $2^{\ell-1} + 2^{\ell-2}$ and consider the polynomial $P(x) = \sum_{T \subseteq [n]:|T| \leq d} \prod_{i \in T} x_i$ in $\mathbb{Z}/8\mathbb{Z}[x_1,\ldots,x_n]$. To prove the theorem, it suffices to show that

$$\Pr_{x \sim \{0,1\}^n} [P(x) = S_{2^\ell,3}(x)] \geq \frac{1}{2} + \frac{1}{16} - o(1).$$

[4] $F_{H,2}$ denotes the 2-lift of F_H.

Clearly, $P(x) = \binom{|x|}{d} \bmod 8$, and

$$P(x) = \begin{cases} 0 & \text{if } 8 \mid \binom{|x|}{d} \\ 4 & \text{if } 4 \mid \binom{|x|}{d} \text{ but } 8 \nmid \binom{|x|}{d} \end{cases} \tag{4}$$

Let $B(x)$ be the number of borrows required when subtracting d from $|x|$. Rewriting (4) in terms of $B(x)$ using Kummer's theorem (See Theorem 3.2), we get

$$P(x) = \begin{cases} 4 & \text{if } B(x) = 2 \\ 0 & \text{if } B(x) \geq 3 \end{cases} \tag{5}$$

We will need the following lemma.

▶ **Lemma 3.7.** *$P(x) = S_{2^\ell,3}(x)$ if*

1. *$|x|_{\ell-2} = 0$, or*
2. *if $(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1,0,0,0)$.*

Proof. Since $d = 2^{\ell-1} + 2^{\ell-2}$, all the bits of d except $d_{\ell-1}$ and $d_{\ell-2}$ are zero. An important observation is that, when subtracting d from $|x|$, no borrows are required by the bits $|x|_i$, $0 \leq i \leq \ell - 3$.

Using the above observation, the reader can verify that when $(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1,0,0,0)$ the number of borrows required is at least 3 i.e. $B(x) \geq 3$, which in turn implies that $P(x) = 0$. Since $|x|_\ell = 0$, $S_{2^\ell,3}(x) = 0$. This proves the second part of the lemma.

To prove the first part, suppose $|x|_{\ell-2} = 0$. Since $d_{\ell-1} = d_{\ell-2} = 1$, it follows that both $|x|_{\ell-2}$ and $|x|_{\ell-1}$ will need to borrow when subtracting d from $|x|$. As argued before, no borrows are required by the bits before (i.e. less significant than) $|x|_{\ell-2}$, and thus the total number of borrows required by the bits $|x|_i$, $0 \leq i \leq \ell - 1$, is 2.

Note that the bit $|x|_{\ell-1}$ borrows from $|x|_\ell$. Consider the following case analysis:

- Case $|x|_\ell = 1$: $|x|_\ell$ will not need to borrow since $d_\ell = 0$. In fact, none of the bits after (i.e. more significant than) $|x|_\ell$ will need to borrow, and thus $B(x) = 2$. This implies that $P(x) = 4$. We also have $S_{2^\ell,3}(x) = 4$ and hence $P(x) = S_{2^\ell,3}(x)$.
- Case $|x|_\ell = 0$: $|x|_\ell$ will require a borrow and this means $B(x) \geq 3$. This would imply $P(x) = 0$. Since $|x|_\ell = 0$, it follows that $P(x) = S_{2^\ell,3}(x)$.

This completes the proof. ◀

By Lemma 3.7, we have

$$\Pr[P(x) = S_{2^\ell,3}(x)] \geq \Pr[|x|_{\ell-2} = 0] + \Pr\left[(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1,0,0,0)\right] \tag{6}$$

Using Lemma 2.4 from Section 2, we have

$$\Pr[|x|_{\ell-2} = 0] \geq \frac{1}{2} - o(1)$$
$$\Pr[(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1,0,0,0)] \geq \frac{1}{16} - o(1)$$

which, together with (6), implies

$$\Pr[P(x) = S_{2^\ell,3}(x)] \geq \frac{1}{2} + \frac{1}{16} - o(1).$$ ◀

4 Upper bounds for $\gamma_{d,k}(\mathrm{Maj}_n)$

In this section, we show an upper bound on $\gamma_{d,k}(\mathrm{Maj}_n)$ where Maj_n denotes the Majority function on n bits.[5]

▶ **Theorem 4.1.** *For any $k \geq 1, d \in \mathbb{Z}^+$, $\gamma_{d,k}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{10d}{\sqrt{n}}$.*

The proof of Theorem 4.1 presented below is an adaptation of techniques appearing in a work of Green [7], who proved a similar result on the approximability of the parity function by polynomials over the ring $\mathbb{Z}/p^k\mathbb{Z}$, for prime $p \neq 2$.

We will need some definitions and facts about $\mathcal{P}_{d,k}$.

We use π to denote the unique ring homomorphism from $\mathbb{Z}/2^k\mathbb{Z}$ to $\mathbb{Z}/2\mathbb{Z}$. Its kernel $\pi^{-1}(0) = \{a \in \mathbb{Z}/2^k\mathbb{Z} \mid 2^{k-1}a = 0\}$ is the set of non-invertible elements in $\mathbb{Z}/2^k\mathbb{Z}$.

We call a set $S \subseteq \{0,1\}^n$ *forcing* for $\mathcal{P}_{d,k}$ if any polynomial $P \in \mathcal{P}_{d,k}$ that vanishes over S is forced to take a value in $\pi^{-1}(0)$ at all points $x \in \{0,1\}^n$. Formally,

$$(\forall x \in S \quad P(x) = 0) \Rightarrow (\forall y \in \{0,1\}^n \quad \pi(P(y)) = 0).$$

Define the polynomial $\pi(P) \in \mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$ to be the polynomial obtained by applying the map π to each of the coefficients of P. Since a multilinear polynomial in $\mathbb{Z}/2^k\mathbb{Z}[x_1, \ldots, x_n]$ is the zero polynomial iff it vanishes at all points of $\{0,1\}^n$ (by Lemma 2.5), we see that S is forcing iff $(\forall x \in S \;\; P(x) = 0) \Rightarrow \pi(P) = 0$.

Note that any interpolating set for $\mathcal{P}_{d,k}$ (see Theorem 2 for the definition) is forcing for $\mathcal{P}_{d,k}$, but the converse need not be true.

We now adapt the proof of Lemma 11 in [7] to bound the size of forcing sets for $\mathcal{P}_{d,k}$.

▶ **Lemma 4.2.** *If S is forcing for $\mathcal{P}_{d,k}$, then $|S| \geq |\{0,1\}^n_{\leq d}| = \binom{n}{\leq d}$.*

The proof of the above lemma appears in the full version of the paper.

We now use Lemma 4.2 to prove Theorem 4.1.

Proof of Theorem 4.1. We assume throughout that $1 \leq d \leq \frac{\sqrt{n}}{10}$; otherwise, there is nothing to prove. Let $\mathrm{Maj}_{n,k} : \{0,1\}^n \to \mathbb{Z}/2^k\mathbb{Z}$ be the k-lift of the Maj_n function. Let $P \in \mathcal{P}_{d,k}$ be arbitrary and let $S_P = \{x \in \{0,1\}^n \mid P(x) = \mathrm{Maj}_{n,k}(x)\}$. We want to show that $|S_P| \leq 2^n \cdot (\frac{1}{2} + \frac{10d}{\sqrt{n}})$. We will argue by contradiction. So assume that $|S_P| > 2^n \cdot (\frac{1}{2} + \frac{10d}{\sqrt{n}})$.

Let E_P be the complement of S_P, i.e. the set of points where P makes an error in computing $\mathrm{Maj}_{n,k}$. We have $|E_P| < 2^n(\frac{1}{2} - \frac{10d}{\sqrt{n}})$. We will try to find a degree D (for suitable $D \leq \lfloor n/2 \rfloor$) polynomial Q such that Q vanishes at all points in E_P but has the property that $Q(x)$ is a unit (i.e. $\pi(Q(x)) \neq 0$) for some $x \in \{0,1\}^n$. To be able to do this, we need the fact that E_P is not forcing for $\mathcal{P}_{D,k}$. By Lemma 4.2, if E_P is indeed forcing for $\mathcal{P}_{D,k}$, then

$$\begin{aligned}
|E_P| \geq \sum_{i=0}^{D} \binom{n}{i} &= \left(\sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i}\right) - \sum_{i=D+1}^{\lfloor n/2 \rfloor} \binom{n}{i} \\
&\geq 2^{n-1} - (\lfloor n/2 \rfloor - D) \cdot \binom{n}{\lfloor n/2 \rfloor} \\
&\geq 2^n \cdot \left(\frac{1}{2} - \frac{2(\lfloor n/2 \rfloor - D)}{\sqrt{n}}\right) = 2^n \cdot \left(\frac{1}{2} - \frac{4d}{\sqrt{n}}\right)
\end{aligned}$$

[5] We define the majority function as $\mathrm{Maj}_n(x) = 1$ iff $|x| > n/2$.

where the last equality follows if we choose $D = \lfloor n/2 \rfloor - 2d$. This contradicts our upper bound on the size of $|E_P|$. Hence, E_P cannot be forcing for $\mathcal{P}_{D,k}$. In particular, we can find Q that vanishes on E_P and furthermore, $\pi(Q(x)) \neq 0$ for some $x \in \{0,1\}^n$.

We now claim that $\pi(Q(x_0)) \neq 0$ for some x_0 of Hamming weight $> n/2$. To see this, consider the polynomial $Q_1 = \pi(Q)$. By construction of Q, we know that Q_1 is a non-zero polynomial of degree D. Hence, by Lemma 2.5, Q_1 is non-zero when restricted to the Hamming ball of radius $D < n/2$ around the all 1s vector. In particular, this implies that there is an input x_0 of Hamming weight $> n/2$ where $Q_1(x_0)$ is non-zero and hence $\pi(Q(x_0)) \neq 0$, or equivalently $2^{k-1}Q(x_0) \neq 0$. Fix this x_0 for the remainder of the proof. Note that $x_0 \notin E_P$ since Q vanishes on E_P.

Now, consider the polynomial $R(x) = Q(x) \cdot P(x)$. We first show that $R(x) = 0$ for all x of Hamming weight $\leq n/2$. Consider any x of Hamming weight $\leq n/2$. If $x \in E_P$, then $R(x) = 0$ since $Q(x) = 0$. On the other hand, if $x \notin E_P$, then $P(x) = \mathrm{Maj}_{n,k}(x) = 0$ since x has Hamming weight $\leq n/2$. Thus, R vanishes at all inputs of Hamming weight $\leq n/2$.

Since the degree of R is at most $\deg(Q)+\deg(P) = D+d = (\lfloor n/2 \rfloor - 2d)+d \leq \lfloor n/2 \rfloor - d$ and R vanishes at all inputs of $\{0,1\}^n_{\leq n/2}$, this implies (by Lemma 2.5) that R must be 0 everywhere. However, at x_0, $R(x_0) = Q(x_0)P(x_0) = Q(x_0)\mathrm{Maj}_{n,k}(x_0) = 2^{k-1}Q(x_0) \neq 0$. This yields the desired contradiction. ◀

5 Connection to non-classical polynomials

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denote the one dimensional torus. Observing that the additive structure of $\mathbb{F}_2$ is isomorphic to the additive subgroup $\{0, 1/2\} < \mathbb{T}$, we can think of a Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2$ as a function $F : \mathbb{F}_2^n \to \{0, 1/2\}$, and conversely, a map $F : \mathbb{F}_2^n \to \{0, 1/2\}$ as a Boolean function.

Tao and Ziegler [16] give a characterization of non-classical polynomials as follows:

▶ **Definition 5.1** (Tao and Ziegler [16]). A function $F : \mathbb{F}_2^n \to \mathbb{T}$ is a non-classical polynomial of degree $\leq d$ if and only if it has the following form:

$$F(x_1, \ldots, x_n) = \alpha + \sum_{0 \leq e_1,\ldots,e_n \leq 1, k \geq 1 : \sum_i e_i + (k-1) \leq d} \frac{c_{e_1,\ldots,e_n,k} x_1^{e_1} \ldots x_n^{e_n}}{2^k} \pmod 1$$

Here $\alpha \in \mathbb{T}$, and $c_{e_1,\ldots,e_n,k} \in \{0,1\}$ are uniquely determined. α is called the *shift* of F, and the largest k such that $c_{e_1,\ldots,e_n,k} \neq 0$ for some $(e_1,\ldots,e_n) \in \{0,1\}^n$ is called the *depth* of F.

Since we are interested in the agreement of a non-classical polynomial with Boolean (i.e. $\{0, 1/2\}$-valued) functions, we will only consider polynomials with shift $\alpha = \frac{A}{2^k}$, where k is the depth of the polynomial and $A \in \{0, \ldots, 2^k - 1\}$.

▶ Remark. Classical polynomials are non-classical polynomials with $\alpha \in \{0, 1/2\}$ and depth $= 1$. It is easy to see that every classical polynomial corresponds to a Boolean function. It is also not hard to show that every Boolean function can be represented as a classical polynomial.

The following lemma relates our model to non-classical polynomials (the proof is given in the full version of the paper):

▶ **Lemma 5.2.** *Let F be a Boolean function, and $d, k \in \mathbb{Z}^+$, $d \geq k$.*

1. *If there is a non-classical polynomial P of degree d and depth k satisfying* $\mathrm{agr}(F, P) = \gamma$*, then there is a $P' \in \mathcal{P}_{d,k}$ satisfying* $\mathrm{agr}(F_k, P') = \gamma$*, where F_k is the k-lift of F.*

2. *If there is a $P \in \mathcal{P}_{d,k}$ satisfying* $\mathrm{agr}(F_k, P) = \gamma$*, then there is a non-classical polynomial P' of degree $\leq d + k - 1$ and depth k satisfying* $\mathrm{agr}(F, P') = \gamma$*.*

The first part of Lemma 5.2 implies the following corollary of Theorem 4.1:

▶ **Corollary 5.3.** *Let $F : \mathbb{F}_2^n \to \mathbb{T}$ be a non-classical polynomial of degree d. Then,*

$$\Pr_{x \sim \mathbb{F}_2^n}[\mathrm{Maj}_n(x) = F(x)] \leq \frac{1}{2} + O\left(\frac{d}{\sqrt{n}}\right).$$

This proves a conjecture of Bhowmick and Lovett [3] that non-classical polynomials of degree d do not approximate the Majority function any better than classical polynomials of the same degree.

The following is a consequence of Theorem 2.2 and the first part of Lemma 5.2:

▶ **Corollary 5.4.** *Let $\ell \geq 2$. Then, for every classical polynomial $P : \mathbb{F}_2^n \to \mathbb{T}$ of degree $\leq 2^\ell - 1$,*

$$\Pr_{x \sim \mathbb{F}_2^n}[P(x) = S_{2^\ell}(x)] \leq \frac{1}{2} + o(1).$$

On the other hand, the second part of Lemma 5.2 and Theorem 3.6 imply

▶ **Corollary 5.5.** *For every $\ell \geq 2$, there is a non-classical polynomial $F : \mathbb{F}_2^n \to \mathbb{T}$ of degree $\leq 2^{\ell-1} + 2^{\ell-2} + 2$ and depth 3 such that*

$$\Pr_{x \sim \mathbb{F}_2^n}[F(x) = S_{2^\ell}(x)] \geq \frac{9}{16} - o(1).$$

Noting that $2^{\ell-1} + 2^{\ell-2} + 2 < 2^\ell$ for $\ell \geq 4$, Corollary 5.4 and Corollary 5.5 imply the following:

▶ **Theorem 5.6.** *There is a Boolean function $F : \mathbb{F}_2^n \to \{0, 1/2\}$ and $d \geq 1$, such that for every classical polynomial P of degree at most d, we have*

$$\Pr_{x \sim \mathbb{F}_2^n}[F(x) = P(x)] \leq \frac{1}{2} + o(1),$$

but there is a non-classical polynomial P' of degree $d' \leq d$ satisfying

$$\Pr_{x \sim \mathbb{F}_2^n}[F(x) = P'(x)] \geq \frac{1}{2} + \Omega(1).$$

This provides a counterexample to an informal conjecture of Bhowmick and Lovett [3] that, for any Boolean function F, non-classical polynomials of degree d do not approximate F any better than classical polynomials of the same degree.

Acknowledgments. We would like to thank David Barrington for taking the time to explain Szegedy's result [15] to us, Arkadev Chattopadhyay for referring us to Green's result [7], and Swagato Sanyal for helpful discussions. We are also grateful to the anonymous reviewers for their detailed and helpful comments.

References

1 Amir Abboud, Richard Ryan Williams, and Huacheng Yu. More applications of the polynomial method to algorithm design. In *Proc. 26th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 218–230, 2015. doi:10.1137/1.9781611973730.17.

2 Noga Alon and Richard Beigel. Lower bounds for approximations by low degree polynomials over $\mathbb{Z}_m$. In *Proc. 16th IEEE Conf. on Computational Complexity*, pages 184–187, 2001. doi:10.1109/CCC.2001.933885.

3 Abhishek Bhowmick and Shachar Lovett. Nonclassical polynomials as a barrier to polynomial lower bounds. In *Proc. 30th Computational Complexity Conf.*, pages 72–87, 2015. doi:10.4230/LIPIcs.CCC.2015.72.

4 Klim Efremenko. 3-query locally decodable codes of subexponential length. *SIAM J. Comput.*, 41(6):1694–1703, 2012. (Preliminary version in *41st STOC*, 2009). doi:10.1137/090772721.

5 Parikshit Gopalan. Query-efficient algorithms for polynomial interpolation over composites. *SIAM J. Comput.*, 38(3):1033–1057, 2008. (Preliminary version in *17th SODA*, 2006). doi:10.1137/060661259.

6 Ben Joseph Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contributions to Discrete Mathematics*, 4(2), 2009. URL: http://cdm.ucalgary.ca/cdm/index.php/cdm/article/view/133, arXiv:0711.3191.

7 Frederic Green. A complex-number Fourier technique for lower bounds on the mod-m degree. *Comput. Complexity*, 9(1):16–38, 2000. (Preliminary version in *12th STACS*, 1995). doi:10.1007/PL00001599.

8 Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica*, 20(1):71–86, 2000. doi:10.1007/s004930070032.

9 Adam R. Klivans and Rocco A. Servedio. Learning DNF in time $2^{O(n^{1/3})}$. *J. Comput. Syst. Sci.*, 68(2):303–318, 2004. (Preliminary version in *33rd STOC*, 2001). doi:10.1016/j.jcss.2003.07.007.

10 Donald Ervin Knuth. *Fundamental Algorithms*, volume I of *The Art of Computer Programming*. Addison-Wesley, 3rd edition, 1997.

11 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993. (Preliminary version in *30th FOCS*, 1989). doi:10.1145/174130.174138.

12 Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition [Russian]. *Mathematicheskie Zametki*, 41(4):598–607, 1987. (English translation in *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987). URL: http://mi.mathnet.ru/eng/mz4883, doi:10.1007/BF01137685.

13 Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proc. 19th ACM Symp. on Theory of Computing (STOC)*, pages 77–82, 1987. doi:10.1145/28395.28404.

14 Roman Smolensky. On representations by low-degree polynomials. In *Proc. 34th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 130–138, 1993. doi:10.1109/SFCS.1993.366874.

15 Mario Szegedy. *Algebraic Methods in Lower Bounds for Computational Models with Limited Communication*. PhD thesis, University of Chicago, 1989.

16 Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Ann. Comb.*, 16(1):121–188, 2012. doi:10.1007/s00026-011-0124-3.

17 Ryan Williams. New algorithms and lower bounds for circuits with linear threshold gates. In *Proc. 46th ACM Symp. on Theory of Computing (STOC)*, pages 194–202, 2014. doi:10.1145/2591796.2591858.